

## CLAIMS

1           1.       A method for speaker verification, comprising:  
2           collecting a plurality of data from a speaker, wherein the plurality of data  
3           comprises acoustic data and non-acoustic data;  
4           using the plurality of data to generate a template comprising a first plurality of  
5           parameters;  
6           receiving a real-time identity claim from a claimant;  
7           using a plurality of acoustic data and non-acoustic data from the identity claim  
8           to generate a second plurality of parameters; and  
9           comparing the first plurality of parameters to the second plurality of  
10          parameters to determine whether the claimant is the speaker, wherein the first  
11          plurality of parameters and the second plurality of parameters include at least one  
12          purely non-acoustic parameter, including a non-acoustic glottal shape parameter  
13          parameter derived from averaging multiple glottal cycle waveforms.

1           2.       The method of claim 1, wherein the first plurality of parameters and  
2           the second plurality of parameters each comprise:  
3           a pitch parameter extracted using non-acoustic data;  
4           at least one pitch synchronous spectral coefficient extracted using non-  
5           acoustic data;  
6           pitch synchronous auto-regressive and moving average (ARMA) coefficients  
7           extracted using non-acoustic data.

1           3.       The method of claim 1, wherein generating the template comprises:

2 extracting a parameter from each of multiple repetitions of each of a set of test  
3 sentences by the speaker;  
4 producing multiple feature vectors, each corresponding to a parameter;  
5 selecting one of the multiple feature vectors as a guide vector for dynamic  
6 time warping; and  
7 averaging the multiple feature vectors to produce a resultant feature vector that  
8 is part of the template.

1 4. The method of claim 3, wherein collecting the first plurality of data  
2 comprises the speaker uttering each of a set of test sentences, and wherein subsequent  
3 utterances of the test sentences by the speaker cause the template to be updated.

1 5. The method of claim 3, wherein comparing the first plurality of  
2 parameters to the second plurality of parameters to determine whether the claimant is  
3 the speaker comprises:  
4 using a dynamic warping algorithm to calculate a warping distance between a  
5 feature vector in the template and a corresponding feature vector generated from the  
6 second plurality of parameters; and  
7 determining whether the calculated distance is above or below a  
8 predetermined threshold.

1 6. The method of claim 1, wherein the non-acoustic data comprises an  
2 electromagnetic (EM) signal that characterizes a motion of the speaker's tracheal and  
3 glottal tissues.

1           7.       The method of claim 6, wherein the EM signal is sampled during the  
2 middle of phonation.

1           8.       The method of claim 7, wherein a glottal shape parameter (GSP) is  
2 based on averaged multiple glottal cycle waveforms generated when the speaker  
3 utters a test sentence.

1           9.       The method of claim 8, wherein non-consecutive two-glottal cycle  
2 waveforms are averaged to produce the GSP.

1           10.      The method of claim 2, wherein extracting the ARMA coefficients  
2 comprises ARMA pole-zero modeling of a speech system, including computing the  
3 fast Fourier transform of the acoustic data and the non-acoustic data and solving for a  
4 transfer function, wherein the non-acoustic data comprises input of the modeled  
5 speech system, and the acoustic data comprises output of the modeled speech system.

1           11.      The method of claim 2, wherein extracting the ARMA coefficients  
2 comprises ARMA pole-zero modeling of a speech system using a parametric linear  
3 model.

1           12.      The method of claim 5, wherein using a dynamic warping algorithm,  
2 comprises applying constraints comprising:  
3           a monotonicity constraint;  
4           at least one endpoint constraint;  
5           at least one global path constraint; and

6 at least one local path constraint.

1 13. The method of claim 5, wherein the predetermined threshold is chosen  
2 such that a false acceptance error rate and a false rejection error rate are substantially  
3 equal.

1 14. The method of claim 13, wherein each feature vector generated from  
2 the second plurality of parameters has its own equal error rate (EER) based upon a  
3 corresponding warping distance from a feature vector that is part of the template.

1 15. The method of claim 13, wherein EERs of each feature vector  
2 generated from the second plurality of parameters are combined to generate an overall  
3 EER used to evaluate the speaker verification method.

1 16. A system for speaker verification, comprising:  
2 at least one microphone for collecting acoustic data from a speaker's voice;  
3 at least one sensor for collecting non-acoustic data from movements of the  
4 speaker's body;  
5 at least one processor;  
6 a memory device coupled to the processor, wherein the memory device stores  
7 instructions that when executed cause the processor to generate a template using the  
8 acoustic data and non-acoustic data, wherein the template comprises a first plurality  
9 of parameters, wherein when a claimant speaks an identity claim into the at least one  
10 microphone, the instruction further cause the processor to generate a second plurality  
11 of parameters, and to compare the first plurality of parameters to the second plurality

12 of parameters to determine whether the claimant is the speaker, wherein the first  
13 plurality of parameters and the second plurality of parameters include at least one  
14 purely non-acoustic parameter, including a non-acoustic glottal shape parameter  
15 parameter derived from averaging multiple glottal cycle waveforms.

1 17. The system of claim 16, wherein the first plurality of parameters and  
2 the second plurality of parameters each comprise:  
3 a pitch parameter extracted using non-acoustic data;  
4 at least one pitch synchronous spectral coefficient extracted using non-  
5 acoustic data;  
6 pitch synchronous auto-regressive and moving average (ARMA) coefficients  
7 extracted using non-acoustic data.

1 18. The system of claim 16, wherein generating the template comprises:  
2 extracting a parameter from each of multiple repetitions of each of a set of test  
3 sentences by the speaker;  
4 producing multiple feature vectors, each corresponding to a parameter;  
5 selecting one of the multiple feature vectors as a guide vector for dynamic  
6 time warping; and  
7 averaging the multiple feature vectors to produce a resultant feature vector that  
8 is part of the template.

1 19. The system of claim 18, wherein collecting the first plurality of data  
2 comprises the speaker uttering each of a set of test sentences, and wherein subsequent  
3 utterances of the test sentences by the speaker cause the template to be updated.

1           20.     The system of claim 18, wherein comparing the first plurality of  
2 parameters to the second plurality of parameters to determine whether the claimant is  
3 the speaker comprises:

4           using a dynamic warping algorithm to calculate a warping distance between a  
5 feature vector in the template and a corresponding feature vector generated from the  
6 second plurality of parameters; and

7           determining whether the calculated distance is above or below a  
8 predetermined threshold.

1           21.     The system of claim 16, wherein the non-acoustic data comprises a  
2 electromagnetic (EM) signal that characterizes a motion of the speaker's tracheal and  
3 glottal tissues.

1           22.     The system of claim 21, wherein the EM signal is sampled during the  
2 middle of phonation.

1           23.     The system of claim 22, wherein a glottal shape parameter (GSP) is  
2 based on averaged multiple glottal cycle waveforms generated when the speaker  
3 utters a test sentence.

1           24.     The system of claim 23, wherein non-consecutive two-glottal cycle  
2 waveforms are averaged to produce the GSP.

1           25.     The system of claim 17, wherein extracting the ARMA coefficients  
2 comprises ARMA pole-zero modeling of a speech system, including computing the  
3 fast Fourier transform of the acoustic data and the non-acoustic data and solving for a  
4 transfer function, wherein the non-acoustic data comprises input of the modeled  
5 speech system, and the acoustic data comprises output of the modeled speech system.

1           26.     The system of claim 17, wherein extracting the ARMA coefficients  
2 comprises ARMA pole-zero modeling of a speech system using a parametric linear  
3 model.

1           27.     The system of claim 20, wherein using a dynamic warping algorithm,  
2 comprises applying constraints comprising:  
3           a monotonicity constraint;  
4           at least one endpoint constraint;  
5           at least one global path constraint; and  
6           at least one local path constraint.

1           28.     The system of claim 20, wherein the predetermined threshold is chosen  
2 such that a false acceptance error rate and a false rejection error rate are substantially  
3 equal.

1           29.     The system of claim 28, wherein each feature vector generated from  
2 the second plurality of parameters has its own equal error rate (EER) based upon a  
3 corresponding warping distance from a feature vector that is part of the template.

1           30.     The system of claim 28, wherein EERs of each feature vector  
2     generated from the second plurality of parameters are combined to generate an overall  
3     EER used to evaluate the speaker verification system.

1           31.     An electromagnetic medium, having stored thereon instructions that  
2     when executed, cause a processor to:  
3             collect a plurality of data from a speaker, wherein the plurality of data  
4     comprises acoustic data and non-acoustic data;  
5             use the plurality of data to generate a template comprising a first plurality of  
6     parameters;  
7             receive a real-time identity claim from a claimant;  
8             use a plurality of acoustic data and non-acoustic data from the identity claim  
9     to generate a second plurality of parameters; and  
10            compare the first plurality of parameters to the second plurality of parameters  
11     to determine whether the claimant is the speaker, wherein the first plurality of  
12     parameters and the second plurality of parameters include at least one purely non-  
13     acoustic parameter, including a non-acoustic glottal shape parameter parameter  
14     derived from averaging multiple glottal cycle waveforms.

1           32.     The electromagnetic medium of claim 31, wherein the first plurality of  
2     parameters and the second plurality of parameters each comprise:  
3             a pitch parameter extracted using non-acoustic data;  
4             at least one pitch synchronous spectral coefficient extracted using non-  
5     acoustic data;



6 pitch synchronous auto-regressive and moving average (ARMA) coefficients  
7 extracted using non-acoustic data.

1 33. The electromagnetic medium of claim 31, wherein generating the  
2 template comprises:  
3 extracting a parameter from each of multiple repetitions of each of a set of test  
4 sentences by the speaker;  
5 producing multiple feature vectors, each corresponding to a parameter;  
6 selecting one of the multiple feature vectors as a guide vector for dynamic  
7 time warping; and  
8 averaging the multiple feature vectors to produce a resultant feature vector that  
9 is part of the template.

1 34. The electromagnetic medium of claim 33, wherein collecting the first  
2 plurality of data comprises the speaker uttering each of a set of test sentences, and  
3 wherein subsequent utterances of the test sentences by the speaker cause the template  
4 to be updated.

1 35. The electromagnetic medium of claim 33, wherein comparing the first  
2 plurality of parameters to the second plurality of parameters to determine whether the  
3 claimant is the speaker comprises:  
4 using a dynamic warping algorithm to calculate a warping distance between a  
5 feature vector in the template and a corresponding feature vector generated from the  
6 second plurality of parameters; and

7 determining whether the calculated distance is above or below a  
8 predetermined threshold.

1 36. The electromagnetic medium of claim 31, wherein the non-acoustic  
2 data comprises a electromagnetic (EM) signal that characterizes a motion of the  
3 speaker's tracheal and glottal tissues.

1 37. The electromagnetic medium of claim 36, wherein the EM signal is  
2 sampled during the middle of phonation.

1 38. The electromagnetic medium of claim 37, wherein a glottal shape  
2 parameter (GSP) is based on averaged two-glottal cycle waveforms generated when  
3 the speaker utters a test sentence.

1 39. The electromagnetic medium of claim 38, wherein non-consecutive  
2 two-glottal cycle waveforms are averaged to produce the GSP.

1 40. The electromagnetic medium of claim 32, wherein extracting the  
2 ARMA coefficients comprises ARMA pole-zero modeling of a speech system,  
3 including computing the fast Fourier transform of the acoustic data and the non-  
4 acoustic data and solving for a transfer function, wherein the non-acoustic data  
5 comprises input of the modeled speech system, and the acoustic data comprises output  
6 of the modeled speech system.



2 collecting a plurality of data from a speaker;  
3 using the plurality of data to create a plurality of parameters, comprising a  
4 glottal shape parameter (GSP) derived from sensing motion of the tracheal and glottal  
5 tissues; and  
6 generating multiple feature vectors, each corresponding to one of the plurality  
7 of parameters.

1 47. The method of claim 46, wherein creating the plurality of parameters  
2 comprises extracting a parameter from each of multiple repetitions of each of a set of  
3 test sentences by the speaker.

1 48. The method of claim 46, further comprising averaging the multiple  
2 feature vectors to produce a resultant feature vector.

1 49. The method of claim 4, wherein the GSP is based on average glottal  
2 cycle waveforms.

1 50. The method of claim 49, wherein the non-consecutive two-glottal cycle  
2 waveforms are averaged to produce the GSP.